AD_____

Award Number: W81XWH-04-1-0735

TITLE: Operative Therapy and the Growth of Breast Cancer Micrometastases: Cause and Effect

PRINCIPAL INVESTIGATOR: Susan E. Clare, Ph.D.

CONTRACTING ORGANIZATION: Indiana University
Indianapolis, IN 46202-5167

REPORT DATE: August 2005

TYPE OF REPORT: Annual

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
01-08-2005	Annual	30 Jul 2004 – 29 Jul 2005

4. TITLE AND SUBTITLE
Operative Therapy and the Growth of Breast Cancer Micrometastases: Cause and Effect

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-04-1-0735

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Susan E. Clare, Ph.D.

E-mail: sclare@iupui.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Indiana University
Indianapolis, IN 46202-5167

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Purpose: To test the hypothesis that there are growth factors/cytokines released to promote healing of the wound created by the resection of a primary breast cancer which have the unintended consequence of enabling the growth of micrometastatic foci present at the time of operation.
Objectives: This study was designed as a feasibility study to test whether it is technically possible to reliably assay changes in the low molecular weight serum proteome.
Major Findings: Human xenograft breast tumors were established in 9 of 12 nude mice. Blood samples were obtained from the mice immediately prior to extirpation of the primary breast cancer and then again 24 hours, 48 hours and 7 day post-operatively. In the analysis of the serum of 5 of the mice there were 8685 peptides quantified resulting in 5949 proteins. Of these 5949 proteins 155 were identified with high confidence and 11 proteins had significant changes among the time points as a function of time pre or post-op.

15. SUBJECT TERMS
Metastasis, operation, inflammation

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
U	U	U	UU	45	19b. TELEPHONE NUMBER *(include area code)*

Table of Contents

INTRODUCTION

The American Cancer Society estimates that 212,930 new cases of invasive breast cancer will be diagnosed in 2005(3). During this same year, 40,870 women will die of this disease making it the second leading cause of cancer death for women in the United States(3). It is the metastasis of breast cancer to an essential organ, i.e., liver, lungs and/or brain and the consequent dysfunction of these organs, which kills the breast cancer patient. Therefore, solving the riddle of metastasis is tantamount to conquering the breast cancer we can not prevent. Undetectable micrometastatic foci are hypothesized to be present at the time of the resection of the primary breast cancer. This hypothesis was the justification for the initiation of the adjuvant chemotherapy trials in breast cancer. The results of these and later trials, in turn, provided support for the presence of micrometastatic cells as the administration of systemic therapies decreases relapse rates and improve disease specific survival (4, 5). The hypothesis underpinning all experiments described in this report is that rather than being inert and unaffected by the operative procedure to resect the primary tumor, **micrometastatic foci respond to a change in soluble substances, e.g., growth factors, cytokines, and/or angiogenic factors, released as a consequence of the operation into the circulation.** Based on the literature and studies by our colleagues, we hypothesized moreover that it is growth factors released as a function of wound healing, which result in growth factor/cytokine cascades that promote the transition of quiescent metastatic cells out of G_0 into the cell cycle, which induce a neovasculature and activate proliferation. The definitive hypothesis tested in this project was: There are proteins and/or peptides that are released, synthesized and/eliminated as a consequence of the operative therapy of primary breast cancer. This hypothesis was tested experimentally with the following specific aims.

Specific Aim 1. To determine if there are proteins/peptides released, synthesized or eliminated in response to operative therapy for primary breast cancer;
Specific Aim 2. To identify these proteins/peptides.

BACKGROUND

1. **Does the act of surgically removing a primary breast cancer affect the growth kinetics of micrometastatic foci?**

The only way to answer this question unequivocally would be to conduct a randomized clinical trial in which one arm received no treatment for breast cancer, an absolutely unethical proposition. Consulting historical controls is also problematic. Because the breast is both visible and palpable, lesions can manifest themselves to even the most unsophisticated observer. This and the realization the breast lesions have the possibility of being lethal have resulted in attempts at treatment at every epoch of recorded medical history(6). There is a single source of data on the natural history of breast cancer and that comes from the Middlesex Hospital in London. In 1962, Bloom and Richardson published a landmark paper which they presented survival data of 250 women with untreated breast cancer admitted to the hospital from 1805 when the first case was admitted to 1933(7). 74.4% of these women had stage IV [metastatic] disease at presentation to the hospital and 23.2% had stage III [locally advanced]. Data from their series are presented in Figure 1. In this graph the death-specific hazard rate is displayed as a function of time after onset of the initial symptom (a lump in the breast in 83% of patients). The data graphed show a peak at about the 4^{th} to 5^{th} year followed by a near constant plateau. In comparison, Demicheli *et al* plotted the hazard of recurrence versus time for 1173 women treated at the Milan Cancer Institute between 1964 and 1980 with mastectomy alone and no adjuvant chemotherapy(8). (Figure 2) There is a marked difference in the graph of hazard rates versus time when compared to the Middlesex data: there is a biphasic curve with a peak at three years after operation and a second, smaller peak between approximately 7-9 years. Data for recurrences, both local and regional, also show recurrence to

be a non-continuous function with a biphasic distribution of peaks, the first and larger at approximately 18 months and the second at 60 months(8). Data from the University of Chicago shows a similar pattern in that their follow-up data also revealed two peaks for mortality: one wide peak at 2-4 years and a second, narrow peak at 8 years(1, 9). (Figure 2, Karrison)

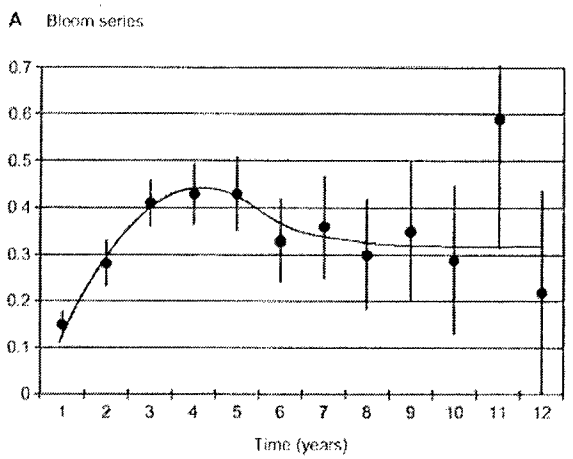


Figure 1 from reference (2)

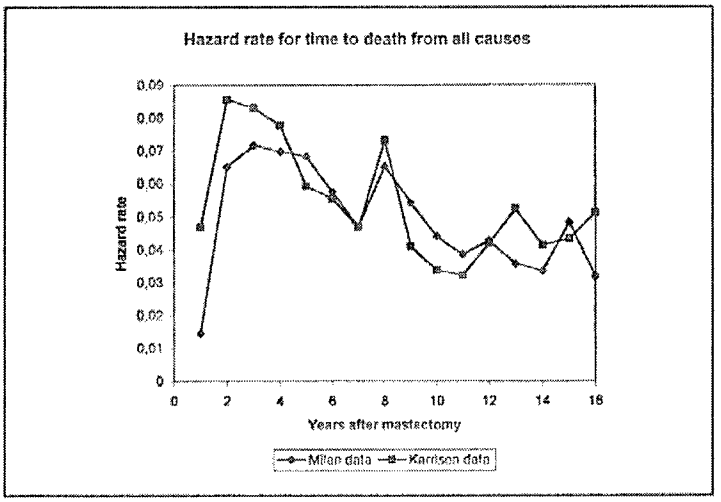


Figure 2 from reference (1)

The growth of tumors, both primary and metastatic has been assumed to follow Gompertzian growth kinetics, that is, near-regular exponential growth at small cell numbers and, decelerated growth at larger cell numbers which is a consequence of the fact that at larger sizes the tumor has outrun its supply of nutrients and oxygen. This continuous growth model would predict that the hazard of recurrence over time should be a continuous function. As indicated above, this is not what is observed in patients who have been treated operatively and therefore, it begs the following questions: Does the act of removing a primary breast cancer have an effect on the growth rate of micrometastatic foci? Is breast tumor growth "non-Gompertzian"? Or both? Is there additional data?

2. Animal studies

a. A study using a transplantable C3H mammary adenocarcinoma, reported in 1979 by Gunduz *et al*, showed that in mice bearing a large and a small tumor focus, removal of the larger tumor focus resulted in an increase in the labeling index and growth fraction, and a decrease in doubling time of the remaining small focus(10). There was a measurable increase in the size of the remaining focus within a week. Since there was minimal change in the DNA synthesis time and in the cell cycle time, the authors concluded that the increase in growth following removal of the larger tumor was probably not the result of a more rapid proliferation of the cells within the cell cycle but, rather, due to recruitment of cells out of G_0. The authors did not postulate the agent or agents responsible for the observed effect on growth kinetics. Numerous other investigators have also inferred that the primary tumor influences the growth rate of metastases or a second tumor transplant(11-16).

b. Studies carried out by Dr. Bernard Fisher and his colleagues in the late 1980s using a variety of mouse tumor models of cancer, C3H, MXT_a, MXT_b, MC54, CD8, and 3LL; revealed that removal of the primary tumor released a factor into the serum which promoted the growth of distant tumor foci(17). Serum obtained less than 18 hours after the extirpation of the primary tumor failed to significantly increase the labeling index of the distant tumor indicating that it is likely that the factor is released in an inactive form.

c. One of the substances released by degranulating platelets at a site of injury, e.g., an operative site, is Transforming Growth Factor β (TGF-β). TGF-β has been implicated in metastatic colorectal carcinoma in the liver (18) and metastatic breast cancer in the bone (19). Kang *et al* used *in vivo* selection in immunodeficient mice to isolate subpopulations of MDA-MB-231 cells highly metastatic to bone(19). MDA-MB-231 is a cell line originally isolated from the plural fluid of a breast cancer patient with a malignant plural effusion. Gene expression profiling identified 43 genes which were over-expressed in the subpopulations highly metastatic to bone and 59 that were under-expressed(http://www.cancercell.org/cgi/content/full/3/6/537/DC1). Many of the over-expressed genes encode cell membrane or secretory proteins which have been implicated in cell homing to bone, angiogenesis, invasion, and osteoclast recruitment. Two of the genes, the cytokine interleukin 11(IL 11), and Connective Tissue Growth Factor (CTGF) are activated by TGF-β. TGF-β is hypothesized to act in a paracrine manner at the site of bony metastasis. Osteolytic resorption releases TGF-β from the bone matrix. TGF-β then stimulates the metastatic breast cells to produce IL 11, and, in addition, Parathyroid Hormone-related Protein (PTHrP) and Vascular Endothelial Growth Factor (VEGF). Unpublished data kindly shared by our colleague at the Indiana University Cancer Research Institute, Dr. Hari Nakshatri, reveals that IL-11 regulates the expression of the chemokine receptor CXCR4. Binding of Stromal Derived Factor-1α (SDF-1α), the only known ligand for CXCR4, increases proliferation (20) and induces chemotaxis and invasion (21) metastatic cells. Thus the cascade is: TGF-β→Il 11→CXCR4:SDF-1α→proliferation and invasion. Although the cascade is hypothesized to occur as a result of TGF-β working in a paracrine fashion, there is no reason to assume that it can not act similarly in an endocrine fashion. It should be noted that a relatively large percentage of platelet derived TGF-β is released into the serum in a latent form, and, therefore, TGF-β is a candidate for the substance identified by Fisher *et al* described in b above.

3. Additional clinical data

a. Additional support for the possibility that operative therapy effects micrometastatic growth comes from the following observation. At presentation, less than 5% of women with T1-T3 breast carcinomas will have evidence of metastatic disease. Nevertheless, by two years following mastectomy, 8% of women with T1 tumors, 20% of women with T2 tumors and 40% of women with T3 tumors will be diagnosed with metastatic breast cancer(22). [Breast cancer is staged according to the size of the tumor (T), the presence or absence of metastatic cells in the lymph nodes (N), and the presence or absence of distant metastasis (M)(23). T1 tumors are ≤ 2 cm in greatest dimension, T2 tumors are >2 cm but ≤ 5 cm, T3 tumors are > 5cm.]

b. Tagliabue *et al* reported that wound drainage fluid as well as serum collected 24 hours after either lumpectomy or mastectomy for invasive breast cancer resulted in an increase in proliferation in all breast cancer cell lines tested in an *in vitro* assay(24). Both wound drainage fluid and serum induced a higher proliferation in HER2-postitive cells than in HER2 negative cells. In an earlier report(25) as well as in this paper, these authors present data that specific members of the Epidermal Growth Factor (EGF) family, such as Heparin Binding-Epidermal Growth Factor (HB-EGF) and Transforming Growth Factor α (TGFα), play a major role in this wound-induced cell proliferation. It should be noted that these factors are also released during degranulation of platelets in the first stage of wound healing. Treatment of the HER2 positive cells with trastuzumab (Herceptin), the humanized monoclonal antibody directed against HER2, before adding the growth stimulus resulted in a significant decrease in drainage-fluid-induced proliferation.

c. Stimulated by the controversy regarding the advisedness of screening mammography for women 40-49 years of age, Retsky *et al* revisited the Milan database referred to above in A(26). A major difference in the early relapse rate was observed when menopausal status was considered. In

premenopausal, node-positive patients, 27% of all distant relapses occurred within the first 10 months following resection. This was twice the rate of the next highest rate, that of postmenopausal, node-positive patients. Retsky and his co-authors use this data to explain the excess mortality seen at years 3-6 for premenopausal women who were screened compared to control in the meta-analysis of six of the mammography screening trials (see Figure 3). That is, using their data they predict that 2-3 years after the onset of screening there will be an excess of 0.11 deaths/1000 screened premenopausal women which they hypothesize is the consequence of operative therapy. The actual patient data shown in Figure 3 reveals a statistically significant increase in mortality in the screened population compared to control of 0.15/1000 at 3-4 years.

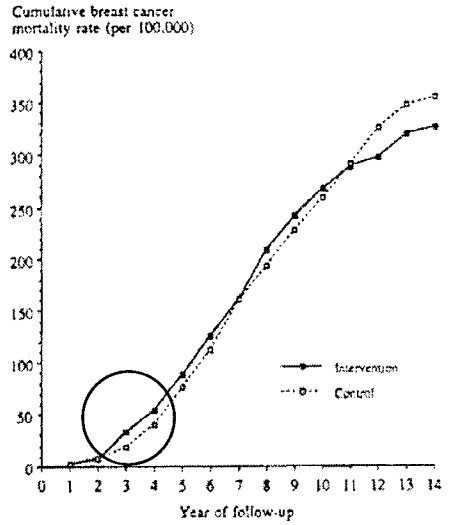


Figure 3 from reference 26

d. A randomized clinical trial comparing open colectomy to laparoscopic-assisted colectomy (LAC) reported that for patients with advanced non-metastatic colon cancer (stage III) LAC was associated with a significantly lower probability of tumor recurrence and a higher probability of overall and cancer-specific survival(27). The improvement in tumor recurrence and survival in these patients was of such a magnitude that these results were similar to those in patients with stage II tumors. The factors responsible for this difference are unknown but one hypothesis is that it is the quantity of growth factors and/or cytokines released as a consequence of the operative intervention that is responsible, a larger amount being released in the more extensive open colectomy.

4. Models to explain bimodal growth pattern

i. Stochastic growth

There are a number of clinical observations of breast cancer growth which can not be explained using Gompertzian kinetics. Speer *et al* observed that the subclinical duration of growth given by the original Gompertz growth equation, is too short (approximately 4 months)(28). Heuser *et al* reported that data derived from serial mammograms indicated that nine out of 109 untreated breast cancers measured over a 1-year period showed no growth; the original Gompertz equation can not account for this observed dormant phase(29). Additionally, breast cancer recurrences are known to occur after a long interval (e.g. 20 years) following primary therapy(30). Thus, Speer *et al* developed a modified Gompertzian model with a stochastic growth rate(28). This allows for a stepwise growth pattern, with the possibility of dormant phases.

With regard to the bimodal recurrence pattern, Retsky *et al* developed a stochastic model to attempt to simulate (by Monte Carlo simulations) the clinical results of the patients enrolled in the Milan database(31). The model consists of a component to describe the growth of the **primary** tumor, based on the model of Speer *et al* discussed above. This is used to describe the release, via a stochastic mechanism, of metastatic cells once the primary tumor is vascularized. The other main component of the model describes **metastatic** growth and detection, and has three main growth stages. The first stage is a dormant single metastatic cell phase. The second is an avascular stage modeled by Gompertzian growth, with a limiting size of approximately 10^5 cells (or about 0.1-0.5 mm in diameter). The size is limited by the fact that the cells must be nourished by diffusion of nutrients from the existing vasculature. Cells in this stage may remain viable but non-growing for an indefinite period of time. Proangiogenic factors or a down-regulation of antiangiogenic factors produced in the stroma, or a combination of both, may result in the induction of a neovasculature that will nourish the metastatic deposit and enable growth. This change accounts for entry into the third stage - a vascular stage that is also modeled by Gompertzian growth with a limiting size of approximately 10^{12} cells. The transition between these three phases is considered stochastic.

One of the interesting features of the model is that it allows for an increased progression of metastatic cells to the avascular and the vascular stages immediately after surgery. The authors hypothesized the increase to be due to a reduction in levels of tumor anti-angiogenesis factor(s) (produced by the primary tumor, e.g., endostatin, angiostatin) after surgery, allowing angiogenesis to occur at the metastasis and thus allowing rapid growth of the metastatic lesion. The model without the possibility of stimulation due to the removal of the primary tumor could not produce the bimodal distribution of relapses observed clinically, and only the second peak was observed. However, when the model included stimulation of the metastatic cells to stage 2 (avascular) or 3 (vascular) due to the surgical removal of the primary tumor, then the bimodal distribution of relapses similar to that seen in the clinical data was observed. Retsky *et al* attributed the first peak of the bimodal recurrence distribution to metastases in the first two stages before surgery that are then promoted to the second or third stages. We point out that the stimulus at the time of surgery does not necessarily need to be the reduction of anti-angiogenesis substances, it is equally probable that it could be pro-angiogenesis factors and/or growth factors and cytokines produced to initiate wound healing. Equally probable is that the substance(s) released propel quiescent micrometastatic foci from G_0 to G_1. The second peak represents cells unperturbed at the time of operative therapy which have undergone a steady stochastic progression from one of the above phases to the next phase.

ii. Stem-cells

Stochastic growth is not the only explanation for the observed clinical results. Al-Hajj and his colleagues at the University of Michigan published a paper detailing their identification of the breast cancer stem-cell(32). Beginning as early progenitor stem-cells, stem-cells mature through a number of stages eventually becoming late differentiated stem-cells. Each of these types of stem-cells or progenitor cells theoretically can be the target of the transforming event(s). Tumors derived from early stem-cells, which by definition are mulitpotential, are hypothesized to lead to a more heterogeneous phenotype then those derived from later, differentiated stem-cells. Tu *et al* have hypothesized that tumors derived from early stem-cells have increased metastatic potential(33). According to this hypothesis, maturation arrest of an early stem-cell would produce a tumor that metastasizes frequently and to various organs, whereas maturation arrest of a later stem cell would produce a tumor which metastasizes rarely if at all. Because early-stem cells maintain their multipotentiality, it is also hypothesized that they have a more diversified growth factor and chemokine receptor profile which may be responsible for the increased metastatic potential. The biphasic curves of the Milan dataset, therefore, may represent two distinct populations of metastatic cells, derived from stem cells at different

stages of maturation. Because of its greater repertoire of receptors, we hypothesize that metastatic foci comprised of early stem-cell derived tumor cells respond to the factor(s) released into the serum at the time of the operation and they are responsible for the first peak. The second peak is a consequence of those cells originating from later stem cells with a more restricted receptor profile which do not contain receptors for the factors elaborated at the time of surgery. Dontu *et al* have carried out gene expression profiling of early progenitor stem-cells and have compared this profile to the one produced from these same cells grown in differentiating conditions(34). A number of receptors have been identified to be up regulated in the early progenitor stem-cells including growth hormone receptor, insulin-like growth factor 2 receptor, fibroblast growth factor receptor 1, Platelet-derived growth factor, β polypeptide.

The two explanations given above are not mutually exclusive in that the soluble factor(s) that the stem-cells respond to may advance cells into the cell cycle or may be angiogenic signals. However, the explanations do differ in that the stochastic model proposes that heterogeneity of phenotype and metastatic potential are the consequence of random, mutational events which occur over time. The stem-cell model argues that the fate of a stem-cell is determined by its state of differentiation at the time it undergoes maturation arrest.

As stated above, the specific aims of this project are:
1.) To determine if there are proteins/peptides released, synthesized or eliminated in response to operative therapy for primary breast cancer;
2.) To identify these proteins/peptides;

Although identification of proteins/peptides is unlikely, by itself, to eliminate one or to establish the validity of the other of the models above, it will provide us with tools to carry out additional hypothesis testing.

BODY
Identification of proteins/peptides

Mouse Xenograft Model:

Dr. Hari Nakshatri and his laboratory colleagues at the Indiana University Cancer Research Institute have developed a mouse model which provided the tool with which to determine if there are proteins/peptides released, synthesized or eliminated in response to operative therapy for primary breast cancer and, if so, what they are. TMD-231 cells were utilized to establish tumors in athymic nu/nu mice. Dr. Nakshatri and his group have determined that these xenograft tumors produce lung metastases **only after** resection of the primary tumor (personal communication).

12 mice were anesthetized with methoxyflurane administered by inhalation. The anterior lateral thorax was prepped with alcohol and betadine to sterilize the skin. A small (8-10 mm) incision was made in the skin over the lateral thorax, allowing visualization of the mammary fat pad. One million TMD-231 cells (in 100 μl of Hepes buffered solution) were injected into the mammary fat pad. The skin was closed with clips and the mouse returned to its cage. Palpable tumors developed in 8 of 12 mice. One mouse died in its cage of unknown cause. Tumor did not develop in the remaining three mice. Tumors were resected after 8 weeks. A blood sample (40 μl) was taken immediately prior to tumor resection and then again 24, 48 and 120 hours after removal of the tumor. [The graphs of labeling index versus time in reference 3 were consulted in order to select these time points.] One mouse expired during the extirpation of the tumor due to exsanguination. Blood was obtained from the facial vein/artery, alternating sides with each subsequent bleed. Blood samples were collected in Sarstedt Microvette 100z capillary tubes. Specimens were centrifuged 1 minute, 2000 RPM (Biofuge fresco, Hearaeus), 4°C to

transfer the blood from the capillary tube to the polypropylene tube to which it was connected. The capillary tubes were then removed and the polypropylene tubes centrifuged 5 minutes, 12,000 RPM (Biofuge fresco, Hearaeus), 4°C to separate the serum. The serum was pipetted into cryovials which were placed into an isopropyl alcohol freezing container which was placed into a -70°C freezer. After all serum samples had been obtained and were frozen, the serum was transferred for storage in liquid nitrogen until transfer to the Proteomics Core Laboratory.

Sample preparation for proteomics:

Albumin and IgG were depleted from the mouse serum samples by Montage™ (Millipore) and Protein G (Amersham) spin columns. The resulting depleted serum samples were denatured by 8M urea, reduced by triethylphosphine, alkylated by iodoethanol, and digested by trypsin(35). This allows all steps to be carried out in one tube without washing or filtering steps.

Protein Identification:

Proteins were identified using two different proteomic methodologies. The first method is called Multi-dimensional Protein Identification Technology (MudPIT). For MudPIT analysis intact serum proteins were proteolytically digested to form a mixture of peptides and were analyzed directly by multidimensional liquid-chromatography, specifically a strong cation exchange column followed by a C-18 reverse phase column, and tandem mass spectrometry. An off-line micro fraction collector was used to collect fractions as they eluted from the cation exchanger. This enabled the use of a continuous salt gradient in the first separation dimension significantly increasing the number of proteins identified in comparison to stepwise elution. Acquired peptide fragmentation spectra were then correlated with predicted amino acid sequences in translated genomic databases using the SEQUEST™ algorithm (Thermo-Finnigan). Specifics of the methodology are provided in the appendices. The advantage of MudPIT is that is enabled the identification of 25-30% more proteins when compared with one dimensional chromatography in a comparative study carried out in our Proteomics Core facility. Its disadvantage is that quantitation is not possible and therefore all data generated is qualitative.

The second method utilized was a label-free single dimension liquid chromatography/mass spectroscopy based quantitative protein analysis. This unique technology combines a proprietary sample preparation protocol(35), liquid chromatography/mass spectroscopy and data analysis tools. It increases the quantifiable protein dynamic range 4- to 5- fold as compared to gel based approaches. Tryptic peptides (~20 μg) were analyzed using a Thermo-Finnigan linear ion-trap mass spectrometer (LTQ) coupled with a HPLC system. A C18 reverse phase column (i.d.=1 mm, length=50 mm) was used to separate peptides with a flow rate of 50 μL/min. Peptides were eluted with a gradient from 5 to 45% acetonitrile developed over 120 min and data were collected in the triple-play mode. Triple play is a Thermo-Finnigan term, meaning: 1) parent ion scan (MS, peptide detection); 2) zoom scan (charge state determination); and 3) MS/MS scan (peptide sequence determination). This system and method can detect at least 1-2 peptides per MS/MS scan. The resulting MS/MS data were applied for database search using SEQUEST™ algorithm (Thermo-Finnigan). Various data processing filters were used to assure that only peptides with the *XCorr* score above 2.0 for singly charged, 2.5 for doubly charged, and 3.8 for triply charged peptides, were analyzed for protein identity. *XCorr* is a cross correlation provided by SEQUEST™ to measure the quality of the peptide identification [the bigger the better]. We were able to obtain quantitative data for up to twenty proteins from each parent ion scan. Using proprietary software(36) and statistical analysis tools, confirmed differentially expressed proteins were identified and the direction of change (up- or down-regulation) was also determined. Also an approximate fold change was calculated, but this is primarily used to determine the significance of the change and not the absolute level of the change (See section below). All data processing were carried out on a Linux cluster

using highly parallel processing and proprietary data qualification and filtering software licensed from Eli Lilly and Company (Indianapolis, IN). Data were then statistically analyzed using multiple proprietary and commercial tools including SAS (See section below).

Protein Quantification: Protein quantification was carried out using the non-gel based and label-free proprietary protein quantification technology that the Core lab has licensed from Eli Lilly and Company. Briefly, once the raw files were acquired from the LTQ, all total ion chromatogram (TIC) were aligned by retention time. Each aligned peak should match parent ion, charge state, daughter ions (MS/MS data) and retention time (within 1 minute window). If any of these parameters were not matched, the peak was thrown out from the quantification. The integral volume under the curve from individually aligned peaks was measured, normalized, and compared for their relative abundance. An example of this quantification process is shown in Fig. 1.

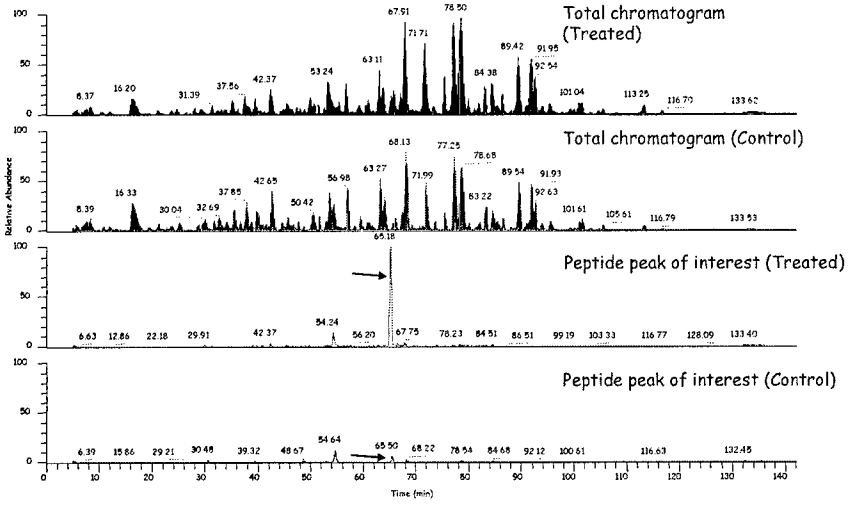


Figure 1. Peptide (protein) quantification by LC/MS.

Statistical Analysis Method Details:

Protein profile comparison of four groups (time points) from five samples (mice) per group. Duplicate injections were performed on an LTQ. There therefore were 4 x 5 x 2 = 40 randomized injections. The data design is as follows:

Table 1.

Group (Time Point)	Number of Samples	Number of Injections
1 (preop)	5	2
2 (24 hrs. post-op)	5	2
3 (48 hrs. post-op)	5	2
4 (7 days post-op)	5	2

Relative expression for each protein discovery is modeled with an ANOVA mixed model:

$Y_{ijk} = \text{mean} + T_i + C_j + TC_{ij} + \text{Injection}(TC)_{ijk}$

- Y = protein expression level (may be arithmetic or log scale)
- T = Treatment group (Time Point) T1 to T4 (effect for T1 to T4)
- C = cell type/animal group C1 to C5 (random effect for cell type/animal C1 through C5)
- TC = interaction effect (e.g. are changes from T1 to T4 the same for each level of C)

- Injection = the random injection effect nested within samples (injections 1-2)

These ANOVA models are called mixed models because there is more than one source of random variation (sample and injection). These models are fit using PROC MIXED in SAS. P-values and Q-values are used to report significant effects. The P-Value is an estimate of the False Positive Rate (FPR) and the Q-value is an estimate of the False Discovery Rate (FDR) which is usually the more relevant of the two error rates. Using terminology from medical diagnostics FPR = 1 – specificity and FDR = 1 – Positive Predictive Value(37).

Group size determination depends on the size of effect to be detected (Fold Change, FC) and the sample-to-sample variation expected (Coefficient of Variation, CV), and which error rates to be controlled. In a situation where we may be testing hundreds of proteins it is best to control the FDR instead of the FPR. The FDR can be large (e.g. > 0.5) even if the FPR is small (e.g. < 0.05). If control of FDR is chosen, then the proportion of proteins that will change (the prevalence) has to be estimated. With this information we can compute the group size required for given power (probability of determining a true change, i.e. the sensitivity). In the following table the Power is fixed at 95%, the FDR is 5%, the FC = 2, and CV = 20%. As the percent of proteins expected to change vary we calculate the group size required.

Table 2. Group Size Determination

% Proteins	Group Size (Power = 95%, FDR = 5%,
5%	4.25
10%	3.81
15%	3.54

Of the 12 mice, sera from 7 were available for analysis. 1 mouse was found dead in its cage on post-op day 28. The cause of death remains unknown. One mouse exsanguinated at the time of tumor extirpation. Three mice did not develop tumors.

Qualitative Analysis: Although historically 60-70% of mice using this model of breast cancer develop lung metastasis, only one mouse in this study developed gross metastatic disease. Therefore, the mice were analyzed as follows:

Sera from the mouse with gross metastatic disease and from one of the mice without metastatic disease were analyzed using MudPIT. As this project had been designed as a feasibility study, MudPIT was chosen as the initial proteomic method as it would yield the maximum number of protein identifications. This analysis cost $27,000 and therefore was limited to the sera from these two mice.

Table 3. Results of MudPIT

sample #	peptides	proteins	highly confident ID
Mouse 5/Bleed 1	23868	3265	367
Mouse 5/Bleed 2	15402	3250	426
Mouse 5/Bleed 3	13661	2796	389
Mouse 5/Bleed 4	13184	2159	248
Mouse 10/Bleed 1	14878	3840	503
Mouse 10/Bleed 2	15349	3765	583
Mouse 10/Bleed 3	16668	3434	498
Mouse 10/Bleed 4	15239	3948	502

In order to identify a protein with "high confidence" two or more unique peptides must be identified.

To reduce the complexity of the data, the proteins for which the identification was highly confident were entered into Pathway Assist (Ariadne Genomics, Inc.) to provide insights in the function of the proteins and their interconnectedness. An example of such a pathway is provided as Figure 2. These are the acute phase proteins identified in Mouse 10 at 48 hours. All proteins for that were identified in our study are encircled in blue.

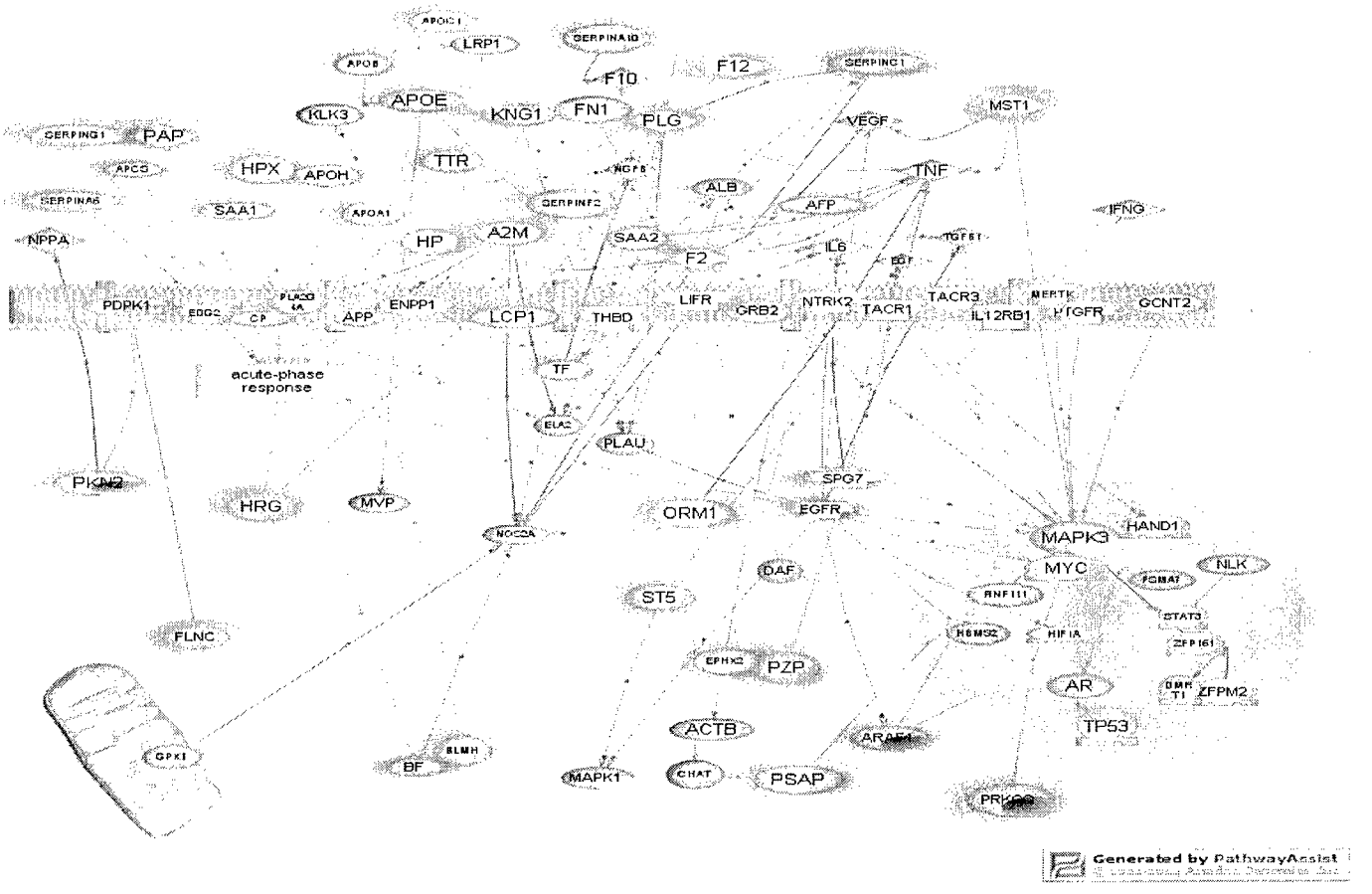


Figure 2. Acute-phase response proteins present at 48 hours in mouse 10

This software analysis has enabled us to hypothesize that there are at least three pathways that may be significantly involved with the proliferation of micrometastatic foci: IL-6 and its receptor gp130, prostaglandins working via their EP4 receptor and the Hepatocyte Growth Factor. Examples of these pathways are provided in the appendices.

Additionally, one of the proteins identified with "high confidence" at 24 hours in Mouse 10 was latent transforming growth factor beta binding protein 3. Latent TGF-beta binding proteins (LTBPs) are required for the proper folding and secretion of TGF-beta. We were unable to quantitate this protein and, therefore, do know if its concentration changes as a function of time. Its presence may provide support for the hypothesis outlined in 2 c above.

Quantitative Analysis:

The same sera as was used for the above qualitative analysis was then analyzed utilizing the label-free single dimension liquid chromatography/mass spectroscopy based quantitative protein analysis. There

were 5109 peptides quantified resulting in 3458 proteins. Of these 3458 proteins there were 102 proteins identified with high confidence (Priority 1). Of the 102 Priority 1 proteins there were 57 that had significant changes as a function of time with respect to the operation. The significance threshold was set to control the False Discovery Rate (FDR) at less than 5%. A False Discovery is a protein declared significant when it isn't. The replicate median % Coefficient of Variation (CV) for the Priority 1 proteins was 4.61%. There were also 1271 proteins that had significant changes among the 3356 proteins that were less confidently identified (Priorities 2-6). The considerations for assigning proteins to the different priority groups are discussed in detail in the statistical summaries in the appendices. Based upon these promising preliminary results, the sera from the remaining 5 mice were analyzed for quantitative differences.

For the remaining 5 mice there were 8685 peptides quantified resulting in 5949 proteins. Of these 5949 proteins there were 155 proteins identified with high confidence (Priority 1). Of the 155 Priority 1 proteins there were 6 that had significant changes among groups (time points). The replicate median % Coefficient of Variation (CV) for the Priority 1 proteins was 6.18% and the combined replicate and sample median %CV was 10.03%. The %CV is the Standard Deviation divided by the Mean on a % scale. There were 5 proteins that had significant changes among the 5794 proteins that were less confidently identified (Priorities 2-6). A sample of the data is presented in table 4.

Table 4. Sample Results of Quantitative Analysis

Rank	Priority	Annotation	Max Fold Change	Significant Change
1	1	similar_to_KIAA1336_protein	1.25728	YES
2	1	Apolipoprotein_A-I_precursor	1.28167	YES
3	1	Zinc_finger_Y-chromosomal_protein_1	1.26396	YES
4	1	_Es1_protein_[Mus_musculus]	1.32484	YES
5	1	Serum_amyloid_A-2_protein_precursor_[Contains:_Amyloid_protein_A_(Amyloid_fibril_protein_AA)]	1.9204	YES
6	1	Apolipoprotein_C-II_precursor	1.23252	YES
7	3	Splice_isoform_1_of_P42703_Leukemia_inhibitory_factor_receptor_precursor	2.14655	YES
8	3	B26300_alpha-1-acid_glycoprotein_(clone_pMAGP3)_-_mouse_(fragment)	1.38323	YES
9	3	Splice_isoform_3_of_O08715_A_kinase_anchor_protein_1,_mitochondrial_precursor	1.58822	YES
10	3	COP9_complex_subunit_6_(COP9	1.58814	YES
11	3	1600031J20Rik_protein	2.73699	YES
12	1	Serum_amyloid_P-component_precursor	1.59702	NO
13	1	Calcium-sensitive_chloride_conductance_protein-1	1.24509	NO
14	1	sex-limited_protein	1.21796	NO
15	1	Transthyretin_precursor	1.52952	NO
16	1	Serotransferrin_precursor	1.17545	NO
17	1	Alpha-1-acid_glycoprotein_1_precursor	1.43212	NO
18	1	Splice_isoform_HMW_of_O08677_Kininogen_precursor_[Contains:_Bradykinin]	1.17502	NO
19	1	Hypothetical_protein	1.1552	NO
20	1	Corticosteroid-binding_globulin_precursor	1.29848	NO
21	1	Alpha-1-acid_glycoprotein_2_precursor	1.38185	NO
22	1	Afamin_precursor	1.09414	NO
23	1	Splice_isoform_1_of_Q01705_Neurogenic_locus_notch_homolog_protein_1_precursor	1.24577	NO
24	1	Serum_amyloid_A-1_protein_precursor	1.87059	NO
25	1	hypothetical_protein_XP_358204	1.18165	NO

KEY RESEARCH ACCOMPLISHMENTS

- Abstract submitted for presentation at the 28th Annual San Antonio Breast Cancer Symposium. Abstract appended.
- Abstract to be submitted for presentation at the Annual Meeting of the Society of Surgical Oncology, March 2006.

REPORTABLE OUTCOMES

- Article regarding this research to be published in the Wall Street Journal September 13, 2005.

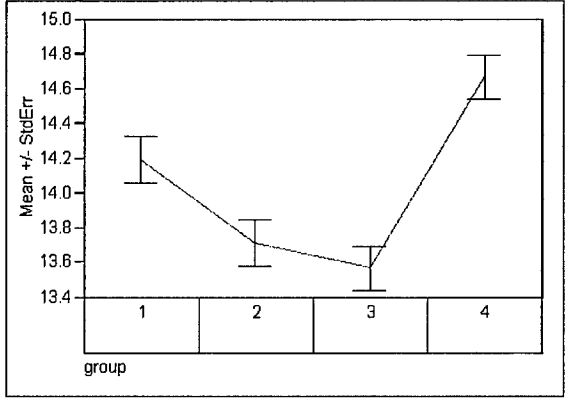
CONCLUSIONS

Using two-dimensional column chromatography and mass spectroscopy (MudPIT) we were able to identify hundreds of serum proteins at each time point prior to and following tumor extirpation. Likewise, using a label-free single dimension liquid chromatography/mass spectroscopy based quantitative protein analysis we were able to identify 5949 proteins of which 155 were identified with a high confidence. 11 of the 5949 proteins had statistically significant changes among the time points relative to tumor extirpation.

This proposal was designed as a feasibility study, that is, to determine if it is technically possible to reliably assay changes in the low molecular weight serum proteome. We have demonstrated that this is technically possible in two different senses: 1.) We have been able to identify hundreds of proteins with high confidence in regard to the reliability of the identification and 2.) The technical variation, i.e., the variation of the measurements, was shown to be relatively small with Coefficients of Variation between injections of 4-6%.

The last set of data was available to us only on the 19th of August and, therefore, data analysis is not yet mature. Nevertheless, we point out that one of the proteins identified with a significant change over the four time points is a splice form of the leukemia inhibitory factor (LIF). LIF has been shown to inhibit the growth of MCF-7 cells(38), and to inhibit the proliferation of non-malignant breast epithelial cells with a reduction in S phase and an increase in cells in G_0/G_1(39). We referred to data from Gunduz *et al* above which indicates that the increase in growth of secondary or metastatic lesions is the result of the recruitment of cells out of G_0(10). We also note that stem cells reside in G_0. Data from our study shows that LIF decreases immediately post-operatively with its nadir at 48 hours as shown in Figure 3 below. From this we can hypothesis that the decrease in LIF may release a brake on proliferation which allows the quiescent micrometastatic cells (stem cells?) to re-enter the cell cycle. We plan to test this hypothesis in the near future.

Figure 3. Change of Leukemia Inhibitory Factor as a Function of Time Pre- and Post-tumor Extirpation



BIBLIOGRAPHY

1. Demicheli, R., Miceli, R., Valagussa, P., and Bonadonna, G. Re: Dormancy of mammary carcinoma after mastectomy. J Natl Cancer Inst, *92:* 347-348., 2000.
2. Demicheli, R., Valagussa, P., and Bonadonna, G. Does surgery modify growth kinetics of breast cancer micrometastases? Br J Cancer, *85:* 490-492., 2001.
3. Jemal, A., Murray, T., Ward, E., Samuels, A., Tiwari, R. C., Ghafoor, A., Feuer, E. J., and Thun, M. J. Cancer statistics, 2005. CA Cancer J Clin, *55:* 10-30., 2005.
4. Polychemotherapy for early breast cancer: an overview of the randomised trials. Early Breast Cancer Trialists' Collaborative Group. Lancet, *352:* 930-942., 1998.
5. Multi-agent chemotherapy for early breast cancer. Cochrane Database Syst Rev CD000487., 2002.
6. De Moulin, D. A Short History of Breast Cancer, p. 138. Dordrecht: Kluwer Academic Publishers, 1989.
7. Bloom, H. J., Richardson, W. W., and Harries, E. J. Natural history of untreated breast cancer (1805-1933). Comparison of untreated and treated cases according to histological grade of malignancy. Br Med J, *5299:* 213-221., 1962.
8. Demicheli, R., Abbattista, A., Miceli, R., Valagussa, P., and Bonadonna, G. Time distribution of the recurrence risk for breast cancer patients undergoing mastectomy: further support about the concept of tumor dormancy. Breast Cancer Res Treat, *41:* 177-185., 1996.
9. Karrison, T. G., Ferguson, D. J., and Meier, P. Dormancy of mammary carcinoma after mastectomy. J Natl Cancer Inst, *91:* 80-85., 1999.
10. Gunduz, N., Fisher, B., and Saffer, E. A. Effect of surgical removal on the growth and kinetics of residual tumor. Cancer Res, *39:* 3861-3865., 1979.
11. Kaplan, H. S. and Murph, E. D. The effect of local roentgen irradiation on the biological behavior of transplantable mouse carcinoma I. Increased frequency of pulmonary metastasis. J Natl Cancer Inst, *9:* 407-413, 1948.
12. Ketcham, A. S., Wexler, H., and Mantel, N. The effect of removal of a "primary" tumor on the development of spontaneous metastases. I. Development of a standardized experimental technic. Cancer Res, *19:* 940-944., 1959.
13. Ketcham, A. S., Kinsey, D. L., Wexler, H., and Mantel, N. The development of spontaneous metastases after the removal of a "primary" tumor. II. Standardization protocol of 5 animal tumors. Cancer, *14:* 875-882., 1961.
14. Sheldon, P. W. and Fowler, J. F. The effect of irradiating a transplanted murine lymphosarcoma on the subsequent development of metastases. Br J Cancer, *28:* 508-514., 1973.
15. Schatten, W. E. An experimental study of postoperative tumor metastases. I. Growth of pulmonary metastases following total removal of primary leg tumor. Cancer, *11:* 455-459., 1958.
16. Van den Brenk, H. A. and Sharpington, C. Effect of local x-irradiation of a primary sarcoma in the rat on dissemination and growth of metastases: dose-response characteristics. Br J Cancer, *25:* 812-830., 1971.
17. Fisher, B., Gunduz, N., Coyle, J., Rudock, C., and Saffer, E. Presence of a growth-stimulating factor in serum following primary tumor removal in mice. Cancer Res, *49:* 1996-2001., 1989.
18. Tsushima, H., Ito, N., Tamura, S., Matsuda, Y., Inada, M., Yabuuchi, I., Imai, Y., Nagashima, R., Misawa, H., Takeda, H., Matsuzawa, Y., and Kawata, S. Circulating transforming growth factor beta 1 as a predictor of liver metastasis after resection in colorectal cancer. Clin Cancer Res, *7:* 1258-1262., 2001.
19. Kang, Y., Siegel, P. M., Shu, W., Drobnjak, M., Kakonen, S. M., Cordon-Cardo, C., Guise, T. A., and Massague, J. A multigenic program mediating breast cancer metastasis to bone. Cancer Cell, *3:* 537-549., 2003.
20. Kijima, T., Maulik, G., Ma, P. C., Tibaldi, E. V., Turner, R. E., Rollins, B., Sattler, M., Johnson, B. E., and Salgia, R. Regulation of cellular proliferation, cytoskeletal function, and signal transduction through CXCR4 and c-Kit in small cell lung cancer cells. Cancer Res, *62:* 6304-6311., 2002.

21. Muller, A., Homey, B., Soto, H., Ge, N., Catron, D., Buchanan, M. E., McClanahan, T., Murphy, E., Yuan, W., Wagner, S. N., Barrera, J. L., Mohar, A., Verastegui, E., and Zlotnik, A. Involvement of chemokine receptors in breast cancer metastasis. Nature, *410:* 50-56., 2001.
22. Koscielny, S., Le, M. G., and Tubiana, M. The natural history of human breast cancer. The relationship between involvement of axillary lymph nodes and the initiation of distant metastases. Br J Cancer, *59:* 775-782., 1989.
23. Singletary, S. E., Allred, C., Ashley, P., Bassett, L. W., Berry, D., Bland, K. I., Borgen, P. I., Clark, G., Edge, S. B., Hayes, D. F., Hughes, L. L., Hutter, R. V., Morrow, M., Page, D. L., Recht, A., Theriault, R. L., Thor, A., Weaver, D. L., Wieand, H. S., and Greene, F. L. Revision of the American Joint Committee on Cancer staging system for breast cancer. J Clin Oncol, *20:* 3628-3636., 2002.
24. Tagliabue, E., Agresti, R., Carcangiu, M. L., Ghirelli, C., Morelli, D., Campiglio, M., Martel, M., Giovanazzi, R., Greco, M., Balsari, A., and Menard, S. Role of HER2 in wound-induced breast carcinoma proliferation. Lancet, *362:* 527-533., 2003.
25. Tagliabue, E., Agresti, R., Ghirelli, C., Morelli, D., and Menard, S. Early relapse of premenopausal patients after surgery for node-positive breast carcinoma. Breast Cancer Res Treat, *72:* 185-186; author reply 186-187., 2002.
26. Retsky, M., Demicheli, R., and Hrushesky, W. Premenopausal status accelerates relapse in node positive breast cancer: hypothesis links angiogenesis, screening controversy. Breast Cancer Res Treat, *65:* 217-224., 2001.
27. Lacy, A. M., Garcia-Valdecasas, J. C., Delgado, S., Castells, A., Taura, P., Pique, J. M., and Visa, J. Laparoscopy-assisted colectomy versus open colectomy for treatment of non-metastatic colon cancer: a randomised trial. Lancet, *359:* 2224-2229., 2002.
28. Speer, J. F., Petrosky, V. E., Retsky, M. W., and Wardwell, R. H. A stochastic numerical model of breast cancer growth that simulates clinical data. Cancer Res, *44:* 4124-4130., 1984.
29. Heuser, L., Spratt, J. S., and Polk, H. C., Jr. Growth rates of primary breast cancers. Cancer, *43:* 1888-1894., 1979.
30. Wheldon, T. E. Mathematical Models in Cancer Research. Bristol and Philadephia: Adam Hilger, 1988.
31. Retsky, M. W., Demicheli, R., Swartzendruber, D. E., Bame, P. D., Wardwell, R. H., Bonadonna, G., Speer, J. F., and Valagussa, P. Computer simulation of a breast cancer metastasis model. Breast Cancer Res Treat, *45:* 193-202., 1997.
32. Al-Hajj, M., Wicha, M. S., Benito-Hernandez, A., Morrison, S. J., and Clarke, M. F. Prospective identification of tumorigenic breast cancer cells. Proc Natl Acad Sci U S A, *100:* 3983-3988., 2003.
33. Tu, S. M., Lin, S. H., and Logothetis, C. J. Stem-cell origin of metastasis and heterogeneity in solid tumours. Lancet Oncol, *3:* 508-513., 2002.
34. Dontu, G., Abdallah, W. M., Foley, J. M., Jackson, K. W., Clarke, M. F., Kawamura, M. J., and Wicha, M. S. In vitro propagation and transcriptional profiling of human mammary stem/progenitor cells. Genes Dev, *17:* 1253-1270., 2003.
35. Hale, J. E., Butler, J. P., Gelfanova, V., You, J. S., and Knierman, M. D. A simplified procedure for the reduction and alkylation of cysteine residues in proteins prior to proteolytic digestion and mass spectral analysis. Anal Biochem, *333:* 174-181., 2004.
36. Higgs, R. E., Knierman, M. D., Gelfanova, V., Butler, J. P., and Hale, J. E. Comprehensive label-free method for the relative quantification of proteins from biological samples. J Proteome Res, *4:* 1442-1450., 2005.
37. Storey, J. D. and Tibshirani, R. Statistical significance for genomewide studies. Proc Natl Acad Sci U S A, *100:* 9440-9445. Epub 2003 Jul 9425., 2003.
38. Douglas, A. M., Goss, G. A., Sutherland, R. L., Hilton, D. J., Berndt, M. C., Nicola, N. A., and Begley, C. G. Expression and function of members of the cytokine receptor superfamily on breast cancer cells. Oncogene, *14:* 661-669., 1997.
39. Grant, S. L., Douglas, A. M., Goss, G. A., and Begley, C. G. Oncostatin M and leukemia inhibitory factor regulate the growth of normal human breast epithelial cells. Growth Factors, *19:* 153-162., 2001.

Appendix 1: MudPIT Methods

Summary Report

Project Number: *010-0016-R*

Multi-dimensional Protein Identification Technology (MudPIT) on Mouse Serum Samples

Prepared for: *Susan Clare, M.D., Ph. D.*

Indiana University School of Medicine

Prepared by: *Mu Wang, Ph.D.*

Jin-Sam (Teddy) You, Ph.D.

Tony Tegeler, Ph.D.

Sheng (Sean) Liu, M.D.

INCAPS

Date: *March 24, 2005*

INCAPS Confidential

351 West 10th Street, Suite 350
Indianapolis, IN 46202
www.indianacaps.com

Summary of the Project

Research Goal:

Profiling proteins in the mouse serum samples
Optimizing serum sample preparation for mass spectrometric analysis (albumin, IgG depletion)
Optimizing LC and nanospray conditions

Materials Received (on 1/25/2005):

Frozen mouse sera, 8 samples (mouse5/Bleed1 to 4 and mouse10/Bleed1 to 4).

Method:

1. Specific information regarding mouse serum sample preparation.

1.1 Thawed mouse serum samples on ice.

1.2 Diluted 2 μL of each sample to 400 μL with d.i. H_2O.

1.3 Determined protein concentration by Bradford assay.

1.4 Diluted 1.25 mg of each sample in 150 μL Montage™ (Millipore) equilibration buffer.

1.5 Added 100 μL of a 25% suspension of G-Sepharose beads in Montage™ equilibration buffer to each sample.

1.6 Incubated samples (shaking or rotating) at RT for approximately 30 minutes.

1.7 Pelleted G-Sepharose beads via centrifugation at 500 x g for 2 minutes.

1.8 Placed 200 μL of the supernatant on a pre-equilibrated Montage™ column.

1.9 Span down Montage™ column (500 x g for 2 minutes); place eluates on the column again spin down Montage™ columns (500 x g for 2 minutes).

1.10 Washed column 2x with 200 μL each of Montage™ wash buffer (500 x g for 2 minutes).

1.11 Placed 120 μL of each sample into individual 1.5-mL Eppendorf tubes.

1.12 Added 120 μL of 8 M urea, 100 mM $(NH_4)_2CO_3$, pH=10.8, 1.25 μg / sample of chicken lysozyme to each sample.

1.13 Added 240 μL of Reduction/Alkylation Cocktail (97.5% ACN, 2% Iodoethanol, 0.5% Triethylphosphine). Incubate at 37 °C for 2 hours.

1.14 Speed-vacuumed overnight.

1.15 Resuspended samples in 200 μL of 100 mM NH_4HCO_3 at pH=8.0.

1.16 Added 400 μL of 100 mM NH_4HCO_3, pH=8.0, 2.5 μg Trypsin solution. Incubate at 37 °C overnight.

1.17 Filtered 100 μL of sample through Millipore Ultrafree MC 0.45 μm filter via centrifugation (500 x g for 2 minutes).

2. Specific information regarding ion-exchange LC sample separation.

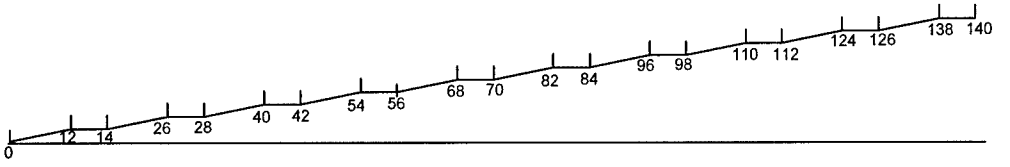
2.1 Packed 5 cm of PolySULFOETHYL (SCX) 300 Å, 5 μm silica into a fritted 200 μm (i.d.) fused silica capillary 10 cm total length of capillary.

2.2 Packed 5 cm of C_{18} 300 Å 5 μm silica into a fritted 200 μm fused silica capillary 10 cm total length of capillary.

2.3 Desalted the sample by loading 12 μg of peptides onto the 5 cm C_{18} column and eluted the peptides onto the SCX column by running a one-hour gradient from 5 to 80% acetonitrile.

2.4 Repeated step 2.3 so a total of 24 μg of peptides were loaded onto the SCX column.

2.5 The chart below illustrates the ammonium acetate salt gradient (4 μL/min) used to elute the peptides from the SCX column. Fractions were taken every 14 minutes resulting in a total of 10 fractions.



14	28	42	56	70	84	98	112	126	140	**min**
25	50	75	100	125	150	175	200	250	500	mM
5%	10%	15%	20%	25%	30%	35%	40%	50%	100%	

3. Specific information regarding sample RP-LC-LTQ-MS/MS.

3.1 Samples were trapped with a C_{18} trapping column and then separated by a reverse phase C_{18} separation column.

3.2 Packed 1 cm of C_{18} 300 Å 5 μm silica into a fritted 200 μm fused silica capillary which were used for the trap column.

3.3 Packed 10 cm of C_{18} 300 Å 5 μm silica into a fritted 100 μm fused silica capillary which were used for the separation column.

3.4 Flow rate was set at 250 μL per minute with a 100 minute gradient from 5 to 80% acetonitrile.

3.5 The LTQ were equipped with a nanospray source for ionization.

3.6 Injected 50 μL from each fraction resulting in 10 LTQ runs for each sample.

4. Data processing and database searching were carried out automatically by the INCAPS Linux cluster.

Summary of Finding

Table 1 below shows the results from MudPIT.

Table 1: MudPIT results from mouse serum after albumin removal using Montage™ kit

Samples	Total Number of Peptides	Total Number of files	Total Number of Proteins Identified	Number of Proteins Identified With High Confidence
Mouse 5 / Bleed 1	23,868	15,506	3,265	367
Mouse 5 / Bleed 2	15,402	9,337	3,250	426
Mouse 5 / Bleed 3	13,661	8,196	2,796	389
Mouse 5 / Bleed 4	13,184	8,094	2,159	248
Mouse 10 / Bleed 1	14,878	9,474	3,480	503
Mouse 10 / Bleed 2	15,349	10,024	3,765	583
Mouse 10 / Bleed 3	16,668	10,864	3,434	498
Mouse 10 / Bleed 4	15,239	9,341	3,948	502

SEQUEST protein identification *XCorr* threshold: singly charged – 1.5; doubly charged – 2.5; triply charged – 3.5. **High confidence**: two or more unique peptides are identified.

Appendix 2: Statistical Summary, Quantitative Method, 2 Mice

Statistical Summary for Project 010-0032-R-01:

Attachments explained in report:
1. **HIGHxcorr-010-0032-R-01.xls (Open with Excel)**
   (summary data on proteins with high identification confidence)
2. **LOWxcorr-010-0032-R-01.xls (Open with Excel)**
   (summary data on proteins with low identification confidence)
3. **VarChart-010-0032-R-01.mht (Open with Internet Explorer)**
   (Plot of intensity levels for proteins)
4. **PeptideIntensities-010-0032-R-01.TXT (Open with text editor)**
   (Normalized and Un-normalized peptide intensities)

Overall Summary:

There were 5109 peptides quantified resulting in 3458 proteins. Of these 3458 proteins there were 102 proteins identified with high confidence (Priority 1). Of the 102 Priority 1 proteins there were 57 that had significant changes* among groups. The significance threshold is set to control the False Discovery Rate (FDR) at less than 5%. A False Discovery is a protein declared significant when it isn't. The replicate median % Coefficient of Variation (CV) for the Priority 1 proteins was 4.61% and the combined replicate and sample median %CV was 12.93%. The %CV is the Standard Deviation divided by the Mean on a % scale. There were also 1271 proteins that had significant changes among the 3356 proteins that were less confidently identified (Priorities 2-6).

***Note:** Because there were only two mice we made Mouse a fixed effect (instead of a random effect) in the statistical model. This means that any significant changes are for these two mice only.

Experimental Design:

The experimental design consisted of 4 time points (groups); 2 samples per time point and 2 replicates per sample:

Group	Time	Number of Samples (Mouse M05 and M10)	Number of Replicates
B1	1	2	2
B2	2	2	2
B3	3	2	2
B4	4	2	2

There was a total of 4 x 2 x 2 = 16 injections randomized to the LTQ. Note that mice M05 and M10 had repeated measures at times 1-4.

May 6, 2005
Kerry Bemis

Proteins Detected:

The proteins quantified are classified according to identification quality as described in the table below:

Xcorr Category	Tier	Priority	Number of Proteins	Median Number of Peptides	Median Number of Sequences	Median XCorr
High	1	1	102	6	5	3.63
	2	2	30	2	1	2.53
	3	3	152	1	1	2.62
High			*284*	*1*	*1*	*2.75*
Low	1	4	216	2	2	2.20
	2	5	172	2	1	2.24
	3	6	2786	1	1	2.16
Low			*3174*	*1*	*1*	*2.17*
			3458	*1*	*1*	*2.18*

Priority assignments reflect our level of confidence in the protein identification. Priority 1 proteins would have the highest likelihood of correct identification and Priority 6 the lowest likelihood of correct identification. This priority system is based on the Xcorr Category and Tier that the protein is assigned. Some experts would view any identification outside of priority 1 as questionable.

Xcorr is a cross correlation provided by Sequest to measure the quality of the peptide identification (the bigger the better). There are various threshold algorithms for deciding which Xcorr values are high quality (High) and which are low quality (Low). I have selected three recommended filters and require a peptide to be high on all three to be scored as High, otherwise it is scored Low. For a protein to be scored High it must have at least one peptide that scored High otherwise it is scored Low. The following table gives the source of the three scoring methods:

Author	Source
You, J.	INCAPS internal communication
Peng, J.	Journal of the Proteome, 2003, Vol. 2, pp. 43-50
Yates, J.	Nat. Biotechnol., 2001, Vol. 19, pp. 242-7

The INCAPS filter results in the following High Xcorr threshold. If a peptide satisfies the rules in the table below it is categorized as High Xcorr; all other peptides are categorized as Low Xcorr. For example all 0 Trypsin cuts are assigned to the Low category.

May 6, 2005
Kerry Bemis

Charge State	Trypsin Cuts	High Xcorr Threshhold
1	2	2.00
2	2	2.50
2	1	3.00
3	2	3.75
3	1	4.00

The assignment of proteins into tiers reflect the number and distinctness of the peptides identified for each protein. Proteins with multiple identified peptides of which at least two have distinct amino acid sequences are classified as Tier 1. Proteins that have multiple identified peptides but differ in ways other than amino acid sequence are classified as Tier 2. For example a protein in Tier 2 could be identified by peptides in two different charge states but the same amino acid sequence (the two different charge states result in different mass/charge ratios resulting in separate identifications and quantifications). Proteins that are identified by a single peptide are classified as Tier 3. These rules are summarized in the table below.

Tier	**Definition**
Tier 1	Proteins with multiple identified amino acid sequences
Tier 2	Proteins with multiple identified peptides (e.g. different charge states)
Tier 3	Proteins with one identified peptide

Protein Quantification:

Every peptide quantified has an intensity measurement for every sample. The intensity measurement is a relative quantity giving the area under the curve (auc) from the select ion chromatogram after background noise removal (auc may be 0). The auc is measured at the same retention time for each sample after the sample chromatograms have been aligned. The intensities are then transformed to the log scale (base 2 is customary) and quantile normalized. Quantile normalization (Bolstad, B. M., et. al., *Bioinformatics*, Vol. 19, No. 2, pp. 185-193) is a method of normalization that essentially ensures that every sample has a peptide intensity histogram of the same scale, location and shape. This normalization removes trends introduced by sample handling, sample preparation, possible total protein differences as well as changes in instrument sensitivity while running multiple samples. If multiple peptides have the same protein identification then their quantile normalized log base 2 intensities are averaged to obtain log base 2 protein intensities. The log base 2 protein intensity is the final quantity that is fit by a separate ANOVA statistical model for each protein. The ANOVA (Analysis of Variance) is a statistical model that separates the variation due to treatments, samples and replicates and constructs the appropriate statistics for discovering treatment differences. The statistical model is covered in more detail at the end of the report.

Summary of Significant Results:

The following table gives the number of proteins with significant changes for each Priority level. The threshold for significance is set to control the **False Discovery Rate (FDR)** for each comparison at 5% (Benjamini, et. al., *Bioinformatics*, 2003, Vol. 19, No. 3, pp. 368-375). The FDR is estimated by the q-value which is an adjusted p-value. The FDR is the proportion of significant changes that are false positives. If proteins with a q-value <= .05 are declared significant it is expected that 5% of the declared changes will be false positives. It is a misconception that the p-value estimates the FDR. The p-Value estimates the False Positive Rate (FPR) which is the proportion of false positives among the proteins that in reality did not change. The FPR = 1- Specificity and FDR = 1 – Positive Predictive Value in the language of medical diagnostics.

The maximum observed absolute Fold Change is also given for each Priority Level.

Fold Change is computed as follows:

Fold Change = Mean Treated Group / Mean Control Group
When Mean Treated Group >= Mean Control Group

Fold Change = - Mean Control Group / Mean Treated Group
When Mean Control Group > Mean Treated Group

Absolute Fold Change = | Fold Change | = absolute or positive value of the Fold Change

A Fold Change of 1 means there is no change.

Also in the table is the Median % Coefficient of Variation (%CV) for each Priority Level. The %CV is the standard deviation / mean on a % scale. The %CV is given both for the injection variation (replicate) as well as the combined injection and sample variation.

Priority	Xcorr Category	Tier	Number of Proteins	Number Significant Changes	Max Absolute Foldchange	Median %CV inj	Median %CV inj + sample
1	High	1	102	57	4.3704781	4.61	12.93
2	High	2	30	15	10.337875	8.91	20.85
3	High	3	152	79	10.672153	8.42	21.32
4	Low	1	216	96	2.2220309	6.94	15.32
5	Low	2	172	81	4.8018379	8.98	17.57
6	Low	3	2786	1000	31.696715	9.45	19.28
			3458	*1328*	*31.696715*	*9.04*	*18.90*

May 6, 2005
Kerry Bemis

The Excel Protein Spread Sheets:
(HIGHxcorr.xls and LOWxcorr.xls)

The High Xcorr and Low Xcorr proteins are listed in separate spreadsheets. The proteins in the spread sheets are ordered by Rank. Rank is assigned by sorting all the proteins in the following order: Priority, significance. Significance is measured by the smallest q-Value among comparisons for a given protein. This means that the proteins are sorted first by priority level (Priority 1-6) and then by significance within priority level. This method of ranking balances both our confidence in the protein identification and our confidence in significant changes. Because all the High Xcorr proteins are in Priority 1-3 and all the Low Xcorr proteins are in Priority 4-6 the High Xcorr spreadsheet has the highest ranking proteins and the Low Xcorr proteins have the lowest ranking proteins. There is a single row for each protein quantified with all the summary information as described below:

Column Name	Description
Rank	Ranked by Priority, MaxConfidence
Annotation	Available Annotation
Max Fold Change	Maximum Absolute Fold Change among the comparisons
SigChange	YES if the Minimum q-Value <= .05 otherwise NO
q_B1_B2, etc.	q-Value comparing group B1 to group B2, etc.
q_Mouse	q-Value comparing Mouse M05 to M10
p_B1_B2, etc.	p-Value comparing group B1 to group B2, etc.
FC_B1_B2, etc.	Fold Change of group B1 relative to group B2, etc.
mean_B1, etc.	The mean protein intensity for group B1, etc.
%CV Rep	% Coefficient of Variation for injection variation
%CV Rep + Sample	% Coefficient of Variation for injection plus sample variation
mean_log2_B1, etc.	The mean of the log base 2 protein intensities for group B1, etc.
se_log2_B1, etc.	The standard error of mean_log2_B1 for group B1, etc.
Protein_id	IPI or NCBI database number
XcorrVal	High or Low as described in the report
Tier	1,2 or 3 as described in the report
Priority	1-6 based on XcorrVal and protein Tier as described in the report
Number_Peptides	The number of distinct identified peptides for this protein
Number_Sequences	The number of distinct amino acid sequences for this protein
mean_xcorr	The mean Xcorr of the peptides identified for this protein

Top_10 Ranked Proteins :

The following table shows the first four columns from the spreadsheet HIGHxcorr.xls. SigChange is YES or NO depending on if the minimum q-value <= .05.

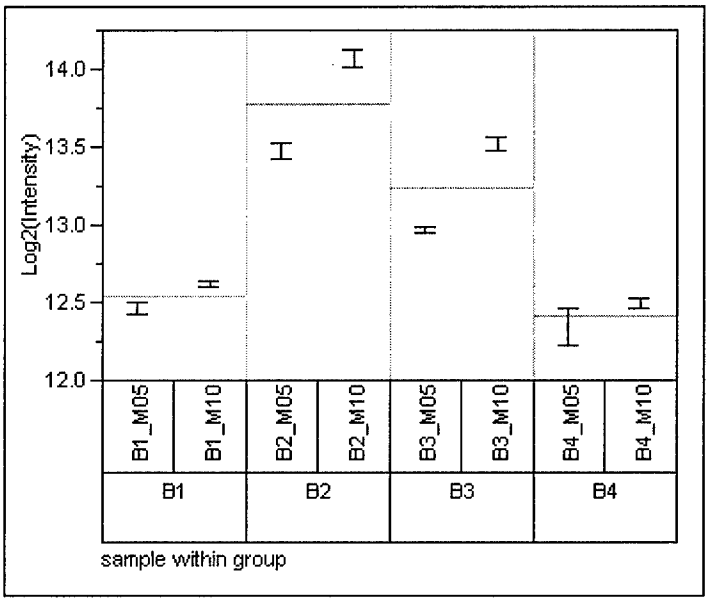
Rank	Annotation	MaxFoldChange	SigChange
1	gi|IPI00118457.1|rs|NP_035444|sp|P05367|Serum_amyloid_A-2_protein_precursor_[Contains:_Amyloid_protein_A_(Amyloid_fibril_protein_AA)]|mass|13622|Mouse	2.56149	YES
2	gi|IPI00139788.2|rs|NP_598738|sp|Q921I1|Serotransferrin_precursor|mass|76724|Mouse	1.3209	YES
3	gi|IPI00114099.2|rs|NP_075603|sp|P33622|Apolipoprotein_C-III_precursor|mass|10982|Mouse	1.6288	YES
4	gi|IPI00127560.1|rs|NP_038725|sp|P07309|Transthyretin_precursor|mass|15776|Mouse	1.70447	YES
5	gi|IPI00267218.2|rs|XP_112207|sp||similar_to_glyceraldehyde-3-phosphate_dehydrogenase_(phosphorylating)_(EC_1.2.1.12)_-_mouse|mass|36621|Mouse	1.50435	YES
6	gi|IPI00340464.2|rs|XP_138241|sp||similar_to_hypothetical_protein_A030003A19|mass|45792|Mouse	1.29726	YES
7	gi|IPI00321666.1|rs|NP_034521|sp|P01898|MHC_(Qa)_Q10-k_class_I_antigen|mass|37251|Mouse	1.23482	YES
8	gi|IPI00378406.1|rs|XP_355197|sp||similar_to_KIAA0944_protein|mass|147610|Mouse	1.66319	YES
9	gi|IPI00319167.2|rs|NP_766504|sp||hydrocephalus_inducing|mass|576642|Mouse	1.68367	YES
10	gi|IPI00323624.1|rs|NP_033908|sp|P01027|Complement_C3_precursor_(HSE-MSF)_[Contains:_C3a_anaphylatoxin]|mass|186483|Mouse	1.30287	YES

Variability charts for the Priority 1 proteins (VarChart.mht):

This file contains separate variability charts for the priority 1 proteins. For each of the proteins there is a plot of the individual protein intensity levels on the log base 2 scale. The horizontal green line is the group mean (i.e. time point mean). Intensities for duplicate injections are joined by a vertical line and each sample is plotted separately. A change in 1 unit on the log base 2 scale represents a doubling or a two fold change. The rank 1 protein plot is shown as an example below. The q-Values comparing each pair of group means are also displayed in the title to the plot. The q-Value estimates the False Discovery Rate (FDR). A brief annotation is also displayed. Detailed information for each protein is in the spread sheets previously described.

May 6, 2005
Kerry Bemis

Rank=1, q_B1_B2=0.0001228976, q_B1_B3=0.0070333297, q_B1_B4=0.6392080505, q_B2_B3=0.1087354557, q_B2_B4=0.0000399594, q_B3_B4=0.0014651868, q_Mouse=0.007955866, Annotation=Serum_amyloid_A-2_protein_precursor_[Contains:_Amyloid_protein_A_(Amyloid_fibril_pr



Peptide Intensities :

This comma delimited text file contains the intensities (auc's) for each peptide as described below:

Column Name	Description
sample_name	Name assigned to this injection
group	Time point
mouse	Mouse number
inj	Replicate injection number
injection_order	Time order when this sample was injected
protein_id	IPI or NCBI database number
peptide_id	Unique peptide identification
zstate	Charge state
num_tryp_cuts	Number of tryptic cuts
xcorr	Sequest cross correlation value
auc	Peptide intensity (area under the curve)
qauc	2^qlog2auc
log2auc	Log2(auc + 1)
qlog2auc	Quantile normalized log2auc

Statistical Methods:

For each protein a separate analysis of variance (ANOVA) model is fit:

Log2(Intensity) = Overall Mean + Time Effect + Mouse Effect + Replicate Effect
(Fixed) (Fixed) (Random)

Because there were only two mice we did not make Mouse a random effect. Any inferences are for these two mice only.

Log2(Intensity) is the protein intensity based on the average of the quantile normalized log base 2 peptide intensities with the same protein identification.

Time Effect refers to the fixed effects (not random) caused by the experimental condition Time that we want to compare.

Mouse Effect refers to the fixed effect due to the two mice.

Replicate Effect refers to the random effects from replicate injections.

All of the injections from one experiment are run in random order on the same LTQ by the same operator.

Because protein intensity is on a log base 2 scale the group means and their differences are converted to arithmetic means and fold change by the following example formulas:

T = Treatment group average of log base 2 protein intensities
C = Control group average of log base 2 protein intensities

We first take antilogs for base 2.

$Mean\_T = 2^T$
$Mean\_C = 2^C$

Finally Fold Change is computed

Fold Change = Mean_T / Mean_C when Mean _T >= Mean_C

Fold Change = - Mean_C / Mean_T when Mean _C > Mean_T

A separate model on the intensity scale was fit to obtain estimates for the replicate and sample + replicate CV's. In this model the two mouse samples were treated as independent for each group and replicates were nested within sample:

Intensity = Overall Mean + Group Effect + Sample Effect + Replicate(Sample) Effect
(Fixed) (Random) (Random)

May 6, 2005
Kerry Bemis

From this model variance components can be estimated for sample and replicate and from this the required CV's can be computed.

When an ANOVA model has two or more random effects it is called a Mixed Model.

The Log2(Intensity) and Intensity models were fit using PROC MIXED in SAS for each protein. The information from the model fit was used to construct the Excel spread sheets.

May 6, 2005
Kerry Bemis

Appendix 3: Statistical Summary, Quantitative Method, 5 Mice

Statistical Summary for Project 010-0048-R2-01-P:

Attachments explained in report:
1. PROTout-010-0048-R2-01-P.xls (Open with Excel)
   (summary data on all individual proteins)
2. SEchart-010-0048-R2-01-P.doc (Open with Word)
   (Plot of mean and standard error for top ranked individual proteins)
3. VARchart-010-0048-R2-01-P.doc (Open with Word)
   (Plot of intensity levels for top ranked individual proteins)

Overall Summary:

There were 8685 peptides quantified resulting in 5949 proteins. Of these 5949 proteins there were 155 proteins identified with high confidence (Priority 1). Of the 155 Priority 1 proteins there were 6 that had significant changes among groups (time points). The significance threshold is set to control the False Discovery Rate (FDR) at less than 5%. A False Discovery is a protein declared significant when it isn't. The replicate median % Coefficient of Variation (CV) for the Priority 1 proteins was 6.18% and the combined replicate and sample median %CV was 10.03%. The %CV is the Standard Deviation divided by the Mean on a % scale. There were 5 proteins that had significant changes among the 5794 proteins that were less confidently identified (Priorities 2-6).

Experimental Design:

The experimental design consisted of 4 groups (time points); 5 samples (mice) per group and 2 replicates per sample:

Group (Time Point)	Number of Samples (Mice)	Number of Replicates
1	5	2
2	5	2
3	5	2
4	5	2

There was a total of 4 x 5 x 2 = 40 injections statistically modeled. The same mouse is sampled at each time point making this a repeated measures design. Originally there were 6 mice but mouse number 1 had sample issues at time points 2 and 3 and was therefore dropped from the statistical analysis.

Proteins Detected:

The proteins quantified are classified according to identification quality as described in the table below:

August 11, 2005, 2005
Kerry Bemis

Xcorr Category	Tier	Priority	Number of Proteins	Median Number of Peptides	Median Number of Sequences	Median XCorr
High	1	1	155	4	3	2.78
	2	2	28	2	1	2.48
	3	3	137	1	1	2.66
Low	1	4	809	2	2	2.21
	2	5	276	2	1	2.30
	3	6	4544	1	1	2.16
Overall			*5949*	*1*	*1*	*2.19*

Priority assignments reflect our level of confidence in the protein identification. Priority 1 proteins would have the highest likelihood of correct identification and Priority 6 the lowest likelihood of correct identification. This priority system is based on the Xcorr Category and Tier that the protein is assigned. Some experts would view any identification outside of priority 1 as questionable.

Xcorr is a cross correlation provided by Sequest to measure the quality of the peptide identification (the bigger the better). There are various threshold algorithms for deciding which Xcorr values are high quality (High) and which are low quality (Low). I have selected three recommended filters and require a peptide to be high on all three to be scored as High, otherwise it is scored Low. For a protein to be scored High it must have at least one peptide that scored High otherwise it is scored Low. The following table gives the source of the three scoring methods:

Author	Source
You, J.	INCAPS internal communication
Peng, J.	Journal of the Proteome, 2003, Vol. 2, pp. 43-50
Yates, J.	Nat. Biotechnol., 2001, Vol. 19, pp. 242-7

The INCAPS filter results in the following High Xcorr threshold. If a peptide satisfies the rules in the table below it is categorized as High Xcorr; all other peptides are categorized as Low Xcorr. For example all 0 Trypsin cuts are assigned to the Low category.

Charge State	Trypsin Cuts	High Xcorr Threshold
1	2	2.00
2	2	2.50
2	1	3.00
3	2	3.75
3	1	4.00

The assignment of proteins into tiers reflect the number and distinctness of the peptides identified for each protein. Proteins with multiple identified peptides of which at least two have distinct amino acid sequences are classified as Tier 1. Proteins that have multiple identified peptides but differ in ways other than amino acid sequence are

classified as Tier 2. For example a protein in Tier 2 could be identified by peptides in two different charge states but the same amino acid sequence (the two different charge states result in different mass/charge ratios resulting in separate identifications and quantifications). Proteins that are identified by a single peptide are classified as Tier 3. These rules are summarized in the table below.

Tier	Definition
Tier 1	Proteins with multiple identified amino acid sequences
Tier 2	Proteins with multiple identified peptides (e.g. different charge states)
Tier 3	Proteins with one identified peptide

Protein Quantification:

Every peptide quantified has an intensity measurement for every sample. The intensity measurement is a relative quantity giving the area under the curve (auc) from the select ion chromatogram after background noise removal (auc may be 0). The auc is measured at the same retention time for each sample after the sample chromatograms have been aligned. The intensities are then transformed to the log scale (base 2 is customary) and quantile normalized. Quantile normalization (Bolstad, B. M., et. al., *Bioinformatics*, Vol. 19, No. 2, pp. 185-193) is a method of normalization that essentially ensures that every sample has a peptide intensity histogram of the same scale, location and shape. This normalization removes trends introduced by sample handling, sample preparation, possible total protein differences as well as changes in instrument sensitivity while running multiple samples. If multiple peptides have the same protein identification then their quantile normalized log base 2 intensities are averaged to obtain log base 2 protein intensities. The log base 2 protein intensity is the final quantity that is fit by a separate ANOVA statistical model for each protein. The ANOVA (Analysis of Variance) is a statistical model that separates the variation due to groups, samples and replicates and constructs the appropriate statistics for discovering group differences. The statistical model is covered in more detail at the end of the report.

Summary of Significant Results:

The following table gives the number of proteins with significant changes for each Priority level. The threshold for significance is set to control the **False Discovery Rate (FDR)** for each comparison and priority at 5% (Benjamini, et. al., *Bioinformatics*, 2003, Vol. 19, No. 3, pp. 368-375). The FDR is estimated by the q-value which is an adjusted p-value. The FDR is the proportion of significant changes that are false positives. If proteins with a q-value <= .05 are declared significant it is expected that 5% of the declared changes will be false positives. It is a misconception that the p-value estimates the FDR. The p-Value estimates the False Positive Rate (FPR) which is the proportion of false positives among the proteins that in reality did not change. The FPR = 1- Specificity and FDR = 1 – Positive Predictive Value in the language of medical diagnostics.

The maximum observed absolute Fold Change is also given for each Priority Level.

Fold Change is computed as follows:

Fold Change = Mean Treated Group / Mean Control Group
When Mean Treated Group >= Mean Control Group

Fold Change = - Mean Control Group / Mean Treated Group
When Mean Control Group > Mean Treated Group

Absolute Fold Change = | Fold Change | = absolute or positive value of the Fold Change

A Fold Change of 1 means there is no change.

Also in the table is the Median % Coefficient of Variation (%CV) for each Priority Level. The %CV is the standard deviation / mean on a % scale. The %CV is given both for the injection variation (replicate) as well as the combined injection and sample variation.

Priority	Xcorr Category	Tier	Number of Proteins	Number Significant Changes	Max Absolute Foldchange	Median %CV inj	Median %CV inj + sample
1	High	1	155	6	2.063265	6.18	10.03
2	High	2	28	0	1.4757084	8.95	11.28
3	High	3	137	5	2.7369903	11.26	13.82
4	Low	1	809	0	1.8841669	8.18	10.51
5	Low	2	276	0	2.1852476	10.11	12.30
6	Low	3	4544	0	4.2039033	11.68	14.28
			5949	*11*	*4.2039033*	*11.02*	*13.48*

August 11, 2005, 2005
Kerry Bemis

The Excel Protein Spread Sheet (PROTout.xls):

The proteins in the spread sheets are ordered by Rank. Rank is assigned by sorting all the proteins in the following order: SigChange(Yes,No), Priority(1-6), Significance. Significance is measured by q_min (the smallest q-Value among comparisons for a given protein). There is a single row for each protein quantified with all the summary information as described below:

Column Name	Description
Rank	Ranked by SigChange, Priority, Significance
Priority	1-6 based on XcorrVal and protein Tier as described in the report
Annotation	Available Annotation
Max Fold Change	Maximum Absolute Fold Change among the comparisons
SigChange	YES if the Minimum q-Value <= .05 otherwise NO
q_min	Minimum q-Value among all two group comparisons
p_min	Minimum p-Value among all two group comparisons
q_1_2, q_1_3	q-Values comparing group 1 to 2, 1 to 3, etc.
p_1_2, p_1_3	p-Values comparing group 1 to 2, 1 to 3, etc.
FC_1_2, etc.	Fold Change of group 2 relative to group 1, etc.
mean_1, mean_2, etc.	The mean protein intensity for group 1, group 2, etc.
%CV Rep	% Coefficient of Variation for injection variation
%CV Rep + Sample	% Coefficient of Variation for injection plus sample variation
mean_log2_1, etc.	The mean of the log base 2 protein intensities for group 1, etc.
se_log2_1, etc.	The standard error of mean_log2_1 for group 1, etc.
Protein_id	IPI or NCBI database number
XcorrVal	High or Low as described in the report
Tier	1,2 or 3 as described in the report
Number_Peptides	The number of distinct identified peptides for this protein
Number_Sequences	The number of distinct amino acid sequences for this protein
mean_xcorr	The mean Xcorr of the peptides identified for this protein

Top 25 Ranked Proteins :

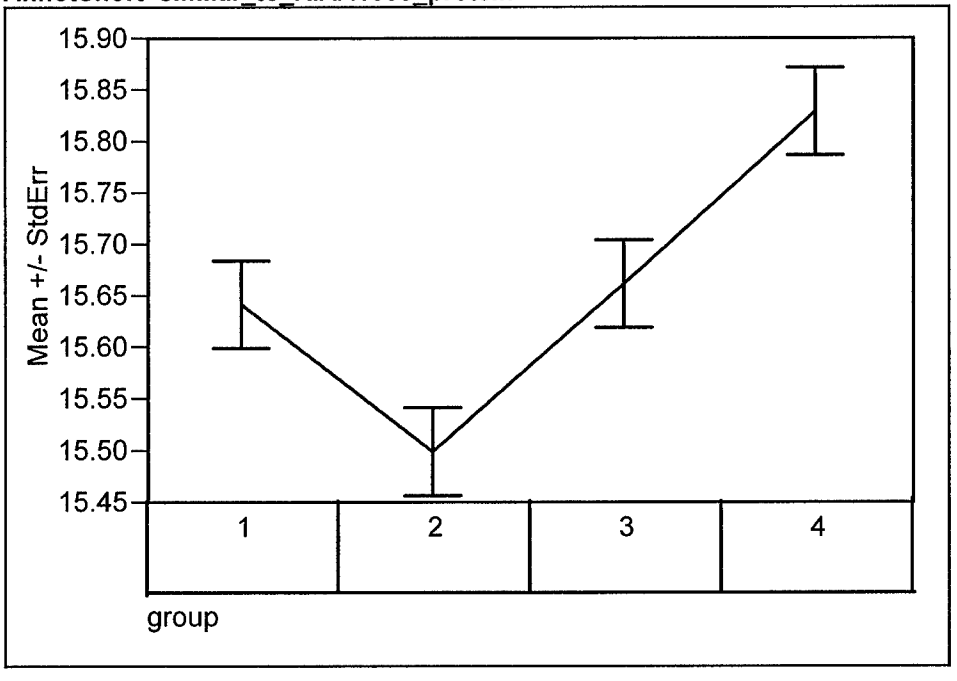
The following table has summary information from the first 5 columns of the first 25 proteins in the Excel spreadsheet described in the last section.

Rank	Priority	Annotation	Max Fold Change	Significant Change
1	1	similar_to_KIAA1336_protein	1.25728	YES
2	1	Apolipoprotein_A-I_precursor	1.28167	YES
3	1	Zinc_finger_Y-chromosomal_protein_1	1.26396	YES
4	1	_Es1_protein_[Mus_musculus]	1.32484	YES
5	1	Serum_amyloid_A-2_protein_precursor_[Contains:_Amyloid_protein_A_(Amyloid_fibril_protein_AA)]	1.9204	YES
6	1	Apolipoprotein_C-II_precursor	1.23252	YES
7	3	Splice_isoform_1_of_P42703_Leukemia_inhibitory_factor_receptor_precursor	2.14655	YES
8	3	B26300_alpha-1-acid_glycoprotein_(clone_pMAGP3)_-_mouse_(fragment)	1.38323	YES
9	3	Splice_isoform_3_of_O08715_A_kinase_anchor_protein_1,_mitochondrial_precursor	1.58822	YES
10	3	COP9_complex_subunit_6_(COP9	1.58814	YES
11	3	1600031J20Rik_protein	2.73699	YES
12	1	Serum_amyloid_P-component_precursor	1.59702	NO
13	1	Calcium-sensitive_chloride_conductance_protein-1	1.24509	NO
14	1	sex-limited_protein	1.21796	NO
15	1	Transthyretin_precursor	1.52952	NO
16	1	Serotransferrin_precursor	1.17545	NO
17	1	Alpha-1-acid_glycoprotein_1_precursor	1.43212	NO
18	1	Splice_isoform_HMW_of_O08677_Kininogen_precursor_[Contains:_Bradykinin]	1.17502	NO
19	1	Hypothetical_protein	1.1552	NO
20	1	Corticosteroid-binding_globulin_precursor	1.29848	NO
21	1	Alpha-1-acid_glycoprotein_2_precursor	1.38185	NO
22	1	Afamin_precursor	1.09414	NO
23	1	Splice_isoform_1_of_Q01705_Neurogenic_locus_notch_homolog_protein_1_precursor	1.24577	NO
24	1	Serum_amyloid_A-1_protein_precursor	1.87059	NO
25	1	hypothetical_protein_XP_358204	1.18165	NO

Standard Error charts for the top 320 proteins (SEchart.doc):

This file contains Standard Error charts for proteins with ranks 1-320. These are proteins in Priorities 1-3. For each of the proteins there is a plot of the individual protein mean intensity levels on the log base 2 scale plus or minus the standard error. The standard error is computed from the statistical model. The blue line connects the group means and helps to visualize the treatment trend. A change in 1 unit on the log base 2 scale represents a doubling or a two fold change. The rank 1 protein plot is shown as an example below. The q_min is the smallest q-Value comparing any two groups and p_min is the smallest p-Value comparing any two groups. A brief annotation is also displayed. Detailed information on all q-Values for each protein is in the spread sheet previously described.

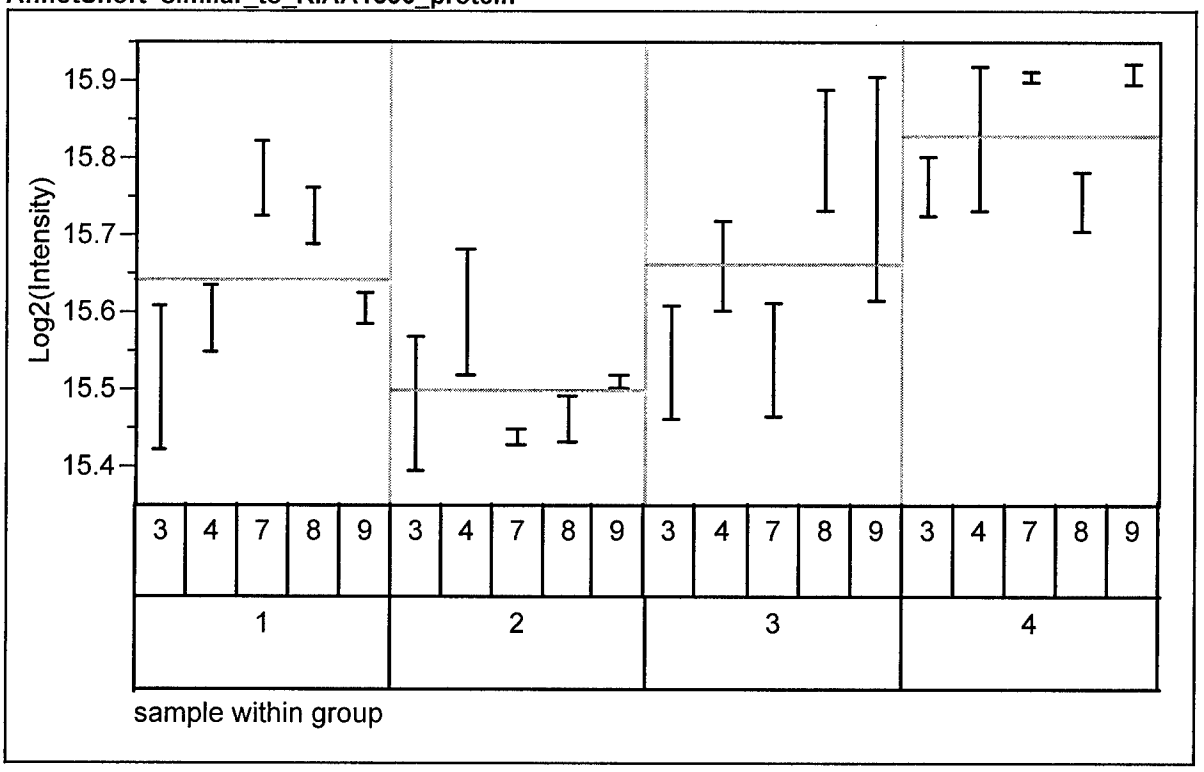
Rank=1, q_min=0.0142097222, p_min=0.0001414353, AnnotShort=similar_to_KIAA1336_protein



Variability charts for the top 320 proteins (VARchart.doc):

This file contains separate variability charts for proteins with ranks 1-320. These are proteins in Priority 1-3. For each of the proteins there is a plot of the individual protein intensity levels on the log base 2 scale. The horizontal green line is the group mean. Intensities for duplicate injections are joined by a vertical line and each sample is plotted separately. A change in 1 unit on the log base 2 scale represents a doubling or a two fold change. The rank 1 protein plot is shown as an example below. The q_min is the smallest q-Value comparing any two groups and p_min is the smallest p-Value comparing any two groups. A brief annotation is also displayed. Detailed information on all q-Values for each protein is in the spread sheet previously described.

Rank=1, q_min=0.0142097222, p_min=0.0001414353, AnnotShort=similar_to_KIAA1336_protein

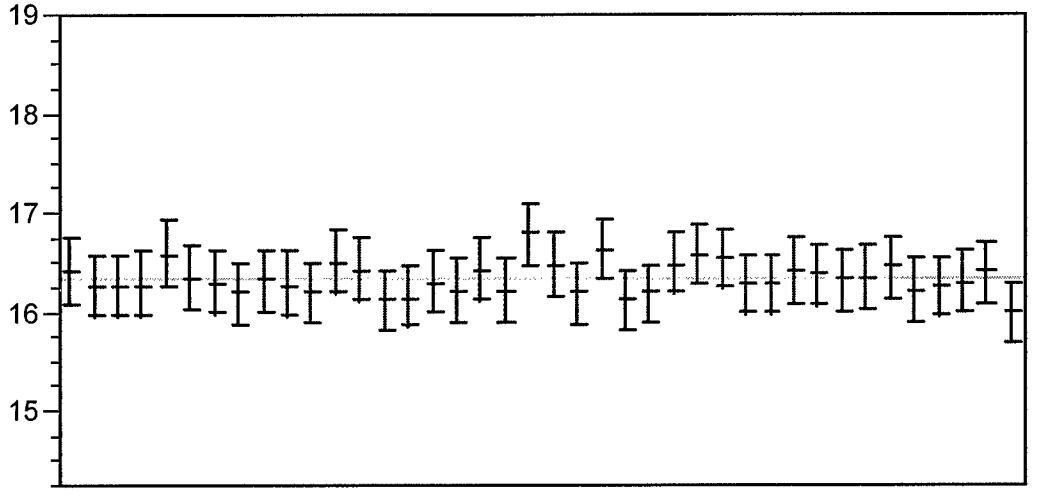


August 11, 2005, 2005
Kerry Bemis .

Spiked Constant Internal Standard:

Chicken lysozyme was spiked at a constant amount of total protein before tryptic digestion. After tryptic digestion 18 peptides were quantified. The plot below is the mean plus and minus the standard error of the 18 peptide log base 2 intensities for each injection. The overall mean is the grey line across the plot. Clearly there is no significant deviation from the overall mean and the quantification is recovering a constant amount of spiked standard for each injection.

Mean Log2(Peptide Intensity) +/- Std Error versus Injection Order



Statistical Methods:

For each protein a separate analysis of variance (ANOVA) model is fit:

Log2(Intensity) =

Overall Mean + Group Effect + Sample Effect + Interaction Effect + Replicate Effect
(Fixed) (Random) (Random) (Random)

Log2(Intensity) is the protein intensity based on the average of the quantile normalized log base 2 peptide intensities with the same protein identification.

August 11, 2005, 2005
Kerry Bemis

Group Effect (time point) refers to the fixed effects (not random) caused by the experimental conditions or treatments that we want to compare including group to group comparisons.

Sample Effect (mouse effect) refers to the random effects from individual biological samples.

Interaction (Time by Mouse) refers the random effects when time trends differ among mice.

Replicate Effect refers to the random effects from replicate injections from the same sample preparation.

All of the injections from one experiment are run in random order on the same LTQ by the same operator.

When an ANOVA model has two or more random effects it is called a Mixed Model.

These models were fit using PROC MIXED in SAS for each protein. The information from the model fit was used to construct the Excel spread sheets.

Because protein intensity is on a log base 2 scale the group means and their differences are converted to arithmetic means and fold change by the following example formulas:

T = Treatment group average of log base 2 protein intensities
C = Control group average of log base 2 protein intensities

We first take antilogs for base 2.

$Mean\_T = 2^T$
$Mean\_C = 2^C$

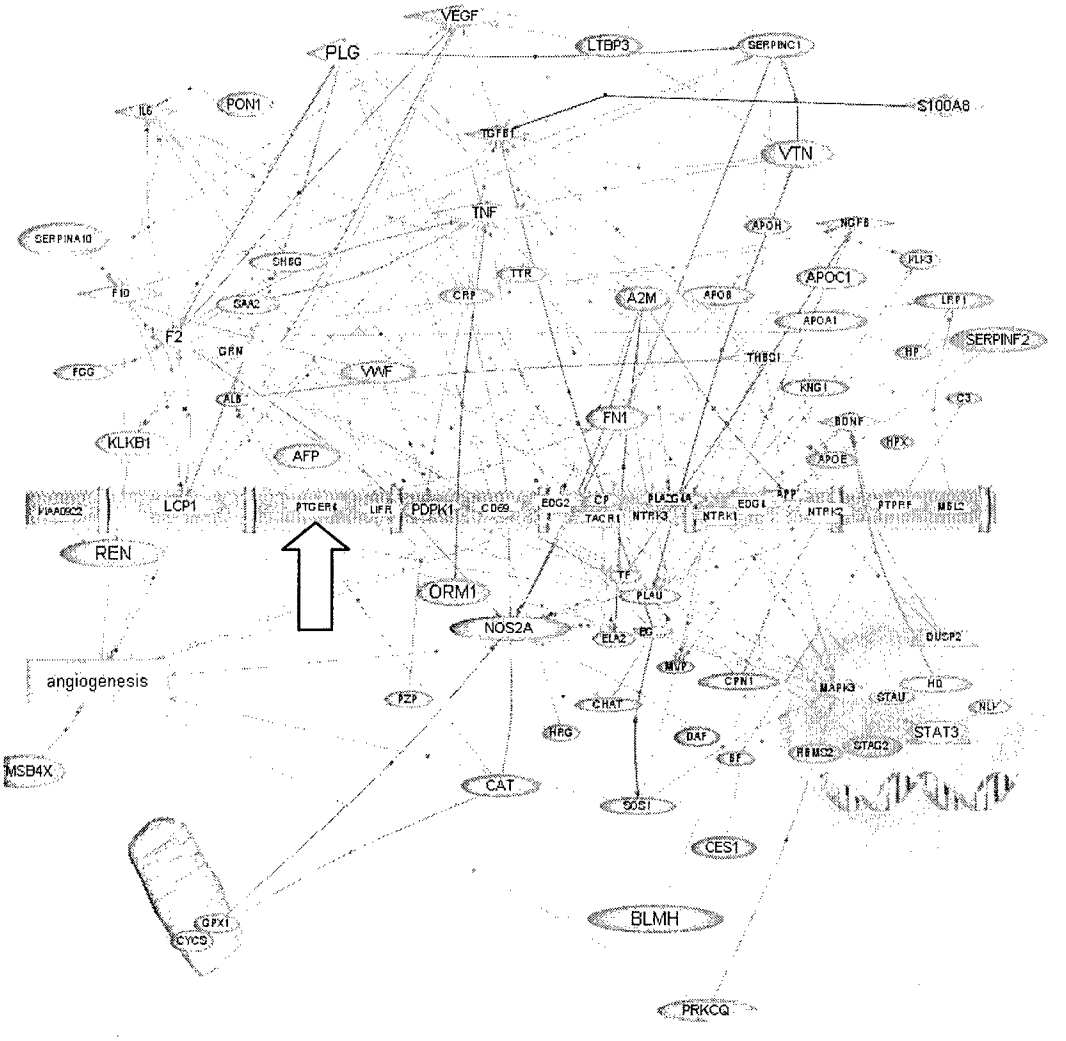
Finally Fold Change is computed

Fold Change = Mean_T / Mean_C when Mean _T >= Mean_C

Fold Change = - Mean_C / Mean_T when Mean _C > Mean_T

Appendix 4

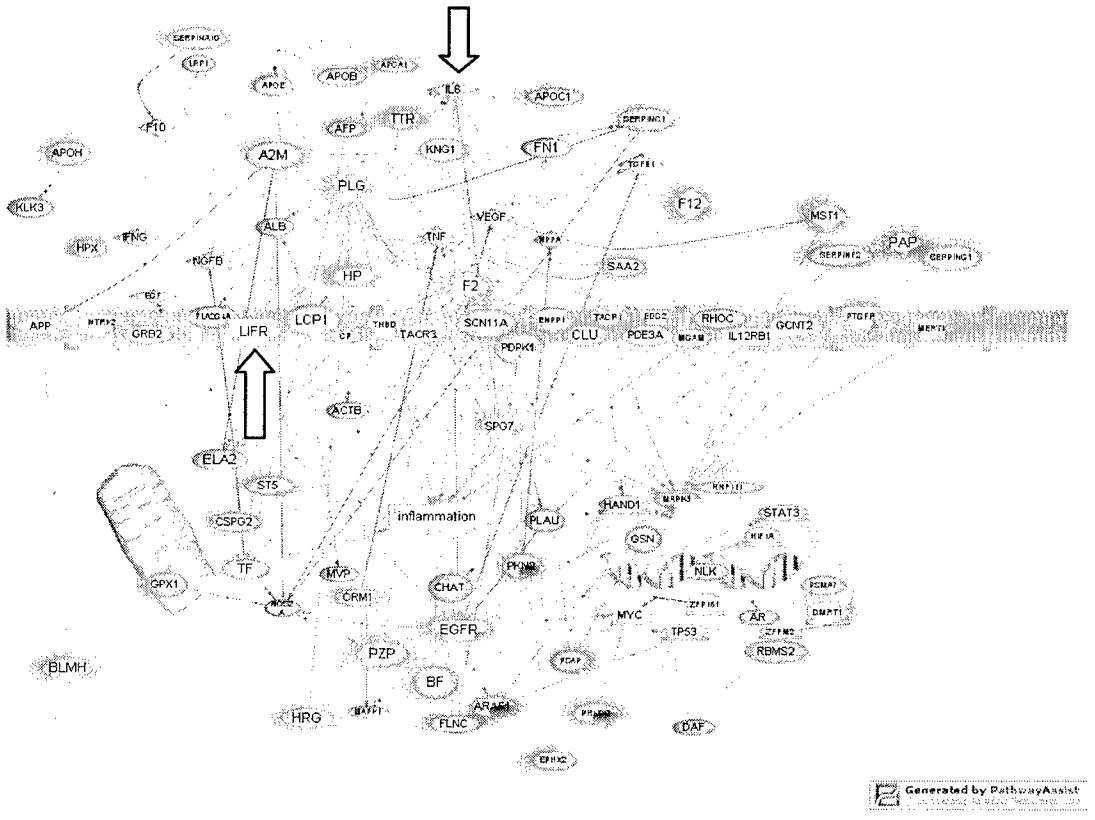
Angiogeneis/Mouse 10/24 hours post-op



Yellow arrow indicates prostaglandin EP4 receptor

Appendix 5

Inflammation/Mouse 10
48 hours post-op



Yellow arrows indicate interleukin 6 (IL 6) and its receptor